

This is a repository copy of *Analysis of User-Generated Comments on Rumor Correction YouTube Videos*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/169103/>

Version: Accepted Version

---

**Proceedings Paper:**

Majid, Gilang Maulana, Pal, Anjan [orcid.org/0000-0001-7203-7126](https://orcid.org/0000-0001-7203-7126), Wardani, Siska Premida et al. (1 more author) (2021) Analysis of User-Generated Comments on Rumor Correction YouTube Videos. In: International Conference on Ubiquitous Information Management and Communication 2021 (proceedings). International Conference on Ubiquitous Information Management and Communication, 04-06 Jan 2021 IEEE .

<https://doi.org/10.1109/IMCOM51814.2021.9377408>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Analysis of User-Generated Comments on Rumor Correction YouTube Videos

Gilang Maulana Majid  
Dept. Southeast Asian Studies  
Goethe University  
Frankfurt, Germany  
gilangmm20@gmail.com

Anjan Pal  
School of Communication and  
Information  
Nanyang Technological University  
Singapore  
anjan001@ntu.edu.sg

Siska Premida Wardani  
Faculty of Economics and  
Business Administration  
Goethe University  
Frankfurt, Germany  
s4204314@stud.uni-frankfurt.de

Snehasish Banerjee  
The York Management School  
University of York  
York, United Kingdom  
snehasish.banerjee@york.ac.uk

**Abstract**—This research investigates how Internet users comment in response to rumor corrections posted on social media. The focus is specifically on the degree to which aggressive language is used. As the test cases for investigation, the research looks into two rumor corrections on YouTube. The rumors were set in the context of the riots and protests in Jakarta following the Indonesian presidential election in 2019. A total of 1,000 comments (500 comments from each of the two cases) was admitted for content analysis. In one case, anti-correction voice was dominant, highlighting the failure of the rumor correction to refute the rumor. In the other, pro-correction voice was dominant, indicating the success of the rumor correction. Aggressive language was widely used in the latter. Implications of the findings are highlighted.

**Keywords**—online hate speech; online rumor; misinformation; rumor correction video; YouTube.

## I. INTRODUCTION

### A. Background

Rumors have been around for centuries [1], and the Internet is only the latest means of communication to be abused for the spread of such unsubstantiated messages [2, 3]. Specifically, communication on social media poses a challenge when rumors spread as if they were facts and become viral quickly on the Internet [4, 5, 6].

Rumors have serious repercussions and sometimes fuel mob violence [7]. For instance, rumors on social media have resulted in fatal mob attacks in countries such as India [7]. Many a time, rumors are propagated deliberately for malicious purposes. The ultimate motives can vary from political to religious [8, 9]. Regardless of the motives, the effectiveness of any rumor campaign directly depends on how much it affects public perceptions in real-time, often referred to as personal involvement [9, 10, 11].

Authorities such as governments, companies, and social media administrators are starting to recognize that rumors must be proactively corrected [12]. Therefore, they are now setting up services (e.g., rumor verification and correction websites, anti-rumor campaigns, etc.) to debunk rumors that are utterly false [13]. Social media channels are also used to spread rumor corrections. For this research, rumor corrections refer to messages that debunk false rumors and provide explanations to strengthen the truthiness of the rebuttals [12, 14, 15].

Although rumor corrections are intended to debunk rumors [14, 15], they might not always work effectively [16, 17]. Rumor corrections sometimes even backfire, which is explained by what is known as the backfire effect—the phenomenon whereby corrective evidence contradicts individuals' prior beliefs and hence ironically strengthens the truthiness of rumors rather than that of rumor corrections [18].

### B. Research Gaps and Research Objectives

Previous works on rumor corrections have mostly examined the psychological factors associated with individuals' worldview [15, 16, 18, 19]. These works suggest that individuals' worldview often influences their processing of the corrections. When the corrections are worldview-dissonant, a backfire effect can occur [16, 18]. Such a backfire effect can be found embedded in anti-correction responses [3], which exhibit not only distrust of the authorities but also a rejection of the given correction. An aggressive tone is often used to voice out such responses.

The use of aggressive language in online comments has been studied under different domains, ranging from its relation to anonymity [20] and cyberbullying [21, 22] to its appearance in news website [23]. Previous works found that aggressive comments often use disparaging words towards a target source, which has been defined as name-calling, and thus, represents a form of online incivility and hate speech [24, 25, 26]. Aggressive comments are likely to make their presence felt amid the cat-and-mouse game between rumors and corrections in the realm of social media communication. Users could reflect their attitudinal dispositions either in favor of or against the argument forwarded by rumor corrections. However, no study has specifically revealed the extent to which such aggressive comments tend to appear in response to rumor corrections in the online setting.

To address the research gap, this research investigates how Internet users comment in response to rumor corrections on social media. The focus is specifically on the degree to which aggressive language is used. As the test cases for investigation, the research looks into two rumor corrections on YouTube. The rumors were set in the context of the riots and protests in Jakarta following the Indonesian presidential election in 2019.

The two cases were chosen for two reasons. First, both cases attracted attention on social media websites. Therefore, a substantial volume of comments could be collected for a

meaningful analysis. Second, public reception to each of the selected case was different: the public mostly accepted one rumor correction while the other was not received favorably. Hence, the two cases allow for an interesting cross-case comparison.

This research makes the following contributions. On the theoretical front, it extends the limited literature on rumor corrections [3, 12, 14] by providing insights into the effectiveness of video rumor corrections on social media. One of the few works on online rumors and corrections to be conducted in the context of Indonesia, it also examines users' responses in terms of online aggression. On the practical front, this research can inform practitioners about factors that determine the effectiveness of rumor corrections.

The rest of the paper proceeds as follows. The following section describes the methods employed in this research, which included two subsections. The first subsection provides a description of the two cases. The second subsection includes the process of data collection and data annotation. The results are discussed in Section III. Thereafter, the paper concludes by highlighting its limitations and future directions.

## II. RESEARCH METHODOLOGY

### A. Descriptions of the Two Cases

The first rumor case was a riot video shot in Kampung Bali. A group of police was shown to be beating a seemingly innocent teenager. This video, which made the rounds on Twitter, was accompanied by the narrative that the victim was a young protester who later passed away due to the attack [27]. The incident happened in the aftermath of the Indonesian presidential election. Later, the National Police confirmed the victim to be neither dead nor a teenager [28]. To correct the rumor, the National Police presented the victim, who was accused of collecting stones to hurl at others amid the unrest [28]. The news conference video was used to correct the rumor, and was disseminated on social media websites, including YouTube.

The second rumor case included a photo that became viral on Facebook and WhatsApp. It showed three policemen wearing face coverings. They were being accused of originally coming from China because of their appearance [29]. The narrative accompanying the rumor photo further asserted that these policemen communicated in Chinese [29]. Two days later, the National Police presented the policemen at a news conference in front of the media [30]. They removed their face coverings and spoke in the Indonesian language. They further confirmed that they were Indonesians, specifically from North Sumatra. To correct the rumor, a video of the news conference, which also shows the hoax spreader's arrest [31], was made available on YouTube.

### B. Data Collection and Annotation

Publicly available comments on YouTube were collected for the two rumor cases. YouTube was chosen because it is one of the widely used social media websites in Indonesia [32]. Moreover, the two rumor correction YouTube videos attracted substantial public comments on YouTube. The videos were searched by using phrases related to the two cases. After that, a relevance check was done on the searched results.

Ten videos for each rumor correction case were identified to collect comments. The top fifty comments from each video

were considered for analysis. Therefore, 1,000 comments were admitted for data analysis: 500 comments related to the first case (50 comments x 10 videos) plus 500 comments related to the second case (50 comments x 10 videos). These comments were in Indonesian, with only a few in local languages such as Javanese and Sundanese.

Content analysis was employed to analyze the data. This methodological lens helps to annotate and interpret data to form themes. Data annotation was done using two annotators who were Indonesians and had received training in qualitative research methods. The annotators were familiar with the context of the two chosen cases. Comments were annotated to identify pro-correction voice (i.e., comments that echoed favorable response to the rumor corrections) or anti-correction voice (i.e., comments that echoed unfavorable response to the rumor corrections). Although some comments were related to the chosen cases, they did not fall under the category of either pro- or anti-correction voices. Such comments were annotated as 'Other'.

With respect to the first case, the annotators found 54 comments to echo pro-correction voice and 268 comments to echo anti-correction voice. Besides, 171 comments were annotated as 'Other' (neither pro-correction nor anti-correction but relevant to the case). The remaining seven were annotated as 'Unrelated' (off-topic).

With respect to the second case, 380 comments echoed the pro-correction voice, and 31 comments echoed the anti-correction voice. While 88 comments were annotated as 'Other', one was annotated as 'Unrelated'. Table I summarizes the comment types in response to the rumor correction YouTube videos.

To identify aggressive comments in each of the rumor correction case, the comment types identified as the majority in both cases were further annotated inductively to find which of them contains a name-calling type of aggression. For the purpose of annotation, the use of disparaging words towards a target source was considered aggressive [24].

One of the two earlier annotators annotated all the comments for the use of aggressive language. The results are discussed below. The excerpts provided are in English and have been translated by one of the annotators. The other annotator has verified the accuracy of all the translations.

## III. RESULTS AND DISCUSSION

Two rumor correction cases studied provide different outlooks. One is dominant with an anti-correction voice, signifying the rumor correction's failure to clarify public perception. While the other has a predominantly pro-correction voice, showing that the rumor correction works as expected.

Concerning the first case, more than half of the responses are against the rumor correction. Specifically, of those 268 comments, 74 of them contain name-calling responses. Concerning the second case, 380 comments (or a little over three quarters) are in favor of the rumor correction, and 188 of them call names. Table II presents the total of the comments that call names.

TABLE I. ANALYSIS OF COMMENTS OF RUMOR CORRECTIONS

Case	#Samples / Example	Comment Types (incl. definition)			
		Pro-correction (i.e., in favor of the rumor corrections)	Anti-correction (i.e., against the rumor corrections)	Others (i.e., Neither pro-correction nor anti-correction voices, but related to the two chosen cases)	Unrelated (i.e., off-topic)
First case	500 (100%)	54 (10.80%)	268 (53.60%)	171 (34.20%)	7 (1.40%)
	Example	<i>"People like him use his brain only to provoke society!"</i>	<i>"Lying to the public again! Whoever gets beaten like that will not survive! The media is blind!"</i>	<i>"I wish whoever lies will get to be punished!"</i>	<i>"Defending deprived rights is more honorable than 1000 traitors."</i>
Second case	500 (100%)	380 (76%)	31 (6.20%)	88 (17.60%)	1 (0.20%)
	Example	<i>"Cheers to the Police! Arrest all traitors and hoax spreaders that only want to create disorder!"</i>	<i>"What a mess! Police from another country!"</i>	<i>"Please find who provided the funding. Don't let him run away!"</i>	<i>"The first one to watch, the first one to comment."</i>

TABLE II. TOTAL OF NAME-CALLING COMMENTS

Cases	#Samples (%)	#Name-calling comments (%)
First case	500 (100%)	74 (14.80%)
Second case	500 (100%)	188 (37.60%)

An aggressive comment uses profane language and appears disrespectfully towards a person or group [24]. Such incivility, when being placed in a particular context, may reduce trust [23]. In the context of responses to rumor corrections, such an aggressive tone often reflects individuals' attitudinal predispositions to keep a side of either pro-or-anti-correction voices.

Such aggression is also found in the backfire effect. Being defined by [18] as resistance from the public toward a rumor correction, responses with a backfire effect appear to disadvantage the parties providing the rumor corrections. These parties—often coming from authoritative sources—face greater distrust from the public when corrections fail.

Concerning the first case, the backfire effect occurred as more than a quarter of comments contain name-calling responses to the authorities issuing the rumor correction. These comments showed negative perceptions of the rumor correction presented and attack relevant authoritative sources. The dataset revealed evidence where individuals were unable to update their viewpoint in light of the corrections. One of the comments showing such inability to update one's viewpoint is as follows: *"What did these policemen think? Such brutality cannot be simply said as 'wrong procedure'! They must be held responsible for this"*. Such comment captures how the public is more critical toward the police's action than the issue of the truthiness of the rumor itself. Although authorities strive to convince individuals about corrections, they might not always be effective in the task of refutation. When the source of a correction is doubted for its credibility, individuals' default worldview can override corrections.

Table III shows the details regarding the parties in the first case whose names were being called by the aggressive comments.

TABLE III. NAME-CALLING AGGRESSIVE COMMENTS IN THE FIRST CASE

	Parties whose names were called (#Name-calling comments = 74)		
	Police	Government	Media
#Comments (%)	55 (74.32%)	14 (18.92%)	5 (6.76%)
Example	<i>"The National Police is crueller than the devil! Your place is in hell!"</i>	<i>"People are smart now. A rogue government!"</i>	<i>"&lt;Name of the victim&gt; is the incumbent's supporter. The media is kind of freakish!"</i>

Three institutions were being called in the aggressive comments: the police, the government, and the media. First, 55 out of the 74 comments (74.32%) indicated that online users call out the police as corrupt and untrustworthy. One such example is as follows: *"the police are good at eliminating traces and making up stories!"*. Another comment, *"Police brutality has been shown to the public! Are these people still deserved to be referred to as our protectors?"* shows that people would instead focus on how badly the police force treated the captured protester. In this case, how the police handled the situation overshadows the debate of whether the victim was a boy, as narrated by the rumor.

Second, 14 out of the 74 comments (18.92%) accused the government of lying and hatching conspiracies. For instance, comments such as *"Hoaxes thrive in a country whose government rules arbitrarily"* and *"the regime's soap opera with <name of the victim> as the leading actor"* were found. Meanwhile, another comment, such as *"the incumbent's administration is afraid of being replaced!"* exposes how the political tension during the presidential election renders the difficult handling of the rumor as people's support is divided.

Lastly, five out of the 74 comments (6.76%) blamed the media. Online users accused the media of helping the

government and the police. One such example is as follows: *“both the regime and the media together created hoaxes!”* Other comments such as, *“Does <name of the media> also want to lead public opinion? You are just one of those media that wants to destroy society!”* and *“Mainstream media lies too much!”* further highlight the attack on the media.

Meanwhile, Table IV displays the details regarding the parties in the second case whose names were being called by the aggressive comments. Concerning the second case, name-calling comments were found to be directed to parties other than the authorities. Nearly half of the pro-correction comments fired back to either one of these three elements: the hoax spreader, the supporters of the contending presidential candidate, and the presidential candidate himself.

TABLE IV. NAME-CALLING AGGRESSIVE COMMENTS IN THE SECOND CASE

	Parties whose names were called (#Name-calling comments = 188)		
	Hoax spreader	Supporters of the contender	Presidential candidate
#Comments (%)	144 (76.60%)	28 (14.90%)	16 (8.50%)
Example	<i>“You should be dead! Why do you want to disturb the order?”</i>	<i>“Those who become the supporters of the contender will automatically be stupid! Always spreading hoax!”</i>	<i>“The contender’s family is all Chinese!”</i>

First, 144 out of the 188 comments (76.60%) indicated condemned the action of the hoax spreader. Comments such as *“Your stupidity is just unreasonable!”* and *“What you said is just an excuse! Say that in front of the judge!”* showed the anger of these online users. Another comment, such as *“I just saw the stupidest netizen in the world,”* shows how the hoax spreader’s action fired back to him.

Second, 28 out of the 188 comments (14.90%) attacked the supporters of the contending presidential candidate. The main concern of these 28 comments was that the online users deemed the <name of the presidential candidate> supporters responsible for the political tension created. Two of them read as, *“Stupid <name of the supporter group>! Spreading hoaxes is their only job.”* and *“<Name of the supporter group> don’t feel alive if they haven’t spread a single hoax!”*

Lastly, 16 out of the 188 comments (8.50%) called the name of the contending presidential candidate himself. They brought up Prabowo’s background information instead, such as his Chinese younger brother, his mother, and his genealogy. The comments were such as, *“The Number 02 (referring to Prabowo) is Chinese, why don’t you protest?”* and *“His younger brother has slanted eyes, too!”* Other comments blamed him for the disorder caused by the widespread rumor. One such example is *“this whole mess is because of Prabowo’s conduct.”*

In sum, these comments show how a rumor has the potential to bring up many other tangentially-related variables into the equation regarding its veracity. Especially when it is in a presidential election, such aggressive comments tend to add more fuel to the polarization already created.

As presented earlier, the authorities took the same procedure when correcting the two rumors; that is, by introducing persons involved in the rumor setting in front of the media to debunk the widespread misinformation. However, as discussed earlier, while the second case seems to be able to correct the rumor and enable the authorities to take control of the situation, the first case falls short.

One of the factors behind such a phenomenon is whether or not existing evidence appears to be in favor of the authority responsible for correcting the rumor. In the second case, the existing evidence is in the form of a photo taken in such close proximity. It makes it easier for people to compare the rumor and the correction, mainly when the policemen themselves appear before the media’s camera correcting the rumor.

Meanwhile, as already discussed in [3], the rumor’s evidence in the first case, which is in the form of short video footage, appears to be visually unclear. As a result, online users can freely make assumptions and easily resist the rumor correction released by authoritative sources. Even when the “supposedly” person involved (in this case is the victim experiencing the beating himself) appears in front of the media to testify that he was the one beaten by the police. It underlines how dependent the authorities could be on the evidence being rumored around when it comes to the success of correcting a rumor.

#### IV. CONCLUSION

Correcting rumors in the realm of social media communication attracts responses from the public, whether they would trust the correction presented or keep denying it and believing that the rumor is the truth. Especially in a political event such as presidential elections, such responses can often contain an aggressive tone that attacks any party deemed untrustworthy by the public. This study uncovers that such messages could appear in one form or the other regardless of whether the majority of the responses given to the rumor correction are in favor of or against.

However, this study suggests that even before the authorities release a rumor correction, one could already predict how the public will respond based on the existing evidence. If the evidence being rumored online is in favor of the authorities correcting it, there would be a good chance that more pro-correction responses will follow. On the contrary, if the evidence is not in their favor, backfire responses will be aimed at the authorities—exacerbating the situation.

It is nonetheless useful to issue corrections to rumors with adequate knowledge to ensure that individuals are less likely to be influenced by rumors. A set of recommendations for creating an effective rebuttal are as follows: First, corrections must emphasize the facts to communicate an explanation that helps debunk a rumor belief. With respect to the first case, the authorities should have admitted the wrong conduct of the police force treating the victim first before proceeding with the refutation of the existing misinformation. It would help the public understand the press conference context better since they do not mix their judgment on the truthiness of the rumor with the police brutality. In the second case, the authorities did not need to admit any wrong conduct. They could proceed directly with correcting the rumor of whether the policemen were of Chinese origin or not.

Second, corrections should rely on evidence. In the first case, the existing video footage was recorded from such a far

distance—resulting in a low-quality source of information that is determinant to the rumor correction's success. A meticulous analysis of the video, such as the body shape of the victim, being compared to the presented victim's appearance, might help the authorities better refute the rumor. In the second case, the evidence in the correction was clear enough for the public once the accused policemen were presented before the media, removed their coverings, and spoke in the Indonesian language.

Third, the source of the corrections must be trustworthy and credible. The source of the correction must be deemed to be unbiased and independent. It would make corrections reliable and help gain public trust. Rather than using mainstream media as the platform to disseminate corrections, authorities can use social media channels to spread corrections as a means to tackle online rumors. Finally, the paper suggests that the use of aggressive language should be discouraged among the public, not only in the specific context of rumors but also in general social media use. There is a need to curb online incivility and hate speech.

This study has two limitations. First, it only compares two cases with a stark difference in terms of how the public responds. Future studies should include more cases so that a more definitive pattern can emerge. Second, the discussion on how existing evidence being rumored could influence public responses should be more in-depth, including the transcript and other supporting materials.

#### REFERENCES

- [1] G. W. Allport, and L. Postman, "An analysis of rumor," *Public Opinion Quarterly*, vol. 10, pp. 501-517, 1946.
- [2] J. A. Gibbons, A. F. Lukowski, and W. R. Walker, "Exposure increases the believability of unbelievable news headlines via elaborate cognitive processing," *Media Psychology*, vol. 7, no. 3, pp. 273-300, 2005.
- [3] G. M. Majid, and A. Pal, "Conspiracy and Rumor Correction: Analysis of Social Media Users' Comments," *Proc of the International Conference on Information and Computer Technologies*, IEEE, Mar. 2020, pp. 331-335.
- [4] A. Y. Chua, and S. Banerjee, "Intentions to trust and share online health rumors: An experiment with medical professionals," *Computers in Human Behavior*, vol. 87, pp. 1-9, 2018.
- [5] Y. Liu, X. Jin, and H. Shen, "Towards early identification of online rumors based on long short-term memory networks", *Information Processing & Management*, vol. 56, no. 4, pp. 1457-1467, 2019.
- [6] A. Pal, A. Y. Chua, and D. H. L. Goh, "Debunking rumors on social media: The use of denials," *Computers in Human Behavior*, vol. 96, pp. 110-122, 2019.
- [7] V. Goel, S. Raj, and P. Ravichandran, "How WhatsApp Leads Mobs to Murder in India," Jul. 2018. Retrieved from <https://www.nytimes.com/interactive/2018/07/18/technology/whatsapp-india-killings.html>
- [8] J. Shin, L. Jian, K. Driscoll, and F. Bar, "Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction," *New Media & Society*, vol. 19, no. 8, pp. 1214-1235, 2017.
- [9] H. Huang, "A war of (mis) information: The political effects of rumors and rumor rebuttals in an authoritarian country," *British Journal of Political Science*, vol. 47, no. 2, pp. 283-311, 2017.
- [10] B. E. Weeks, and R. K. Garrett, "Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 US presidential election," *International Journal of Public Opinion Research*, vol. 26, no. 4, pp. 401-422, 2014.
- [11] A. Pal, and S. Banerjee, "Understanding online falsehood from the perspective of social problem," In I. Chiluwu, & S. Samoilenko (Eds.), *Handbook of Research on Deception, Fake News, and Misinformation Online*, pp. 1-17, 2019, IGI Global.
- [12] A. Pal, A. Y. Chua, and D. H. L. Goh, "How do users respond to online rumor rebuttals?," *Computers in Human Behavior*, vol. 106, 106243, 2020.
- [13] A. Y. Chua, and S. Banerjee, "Rumor verifications on Facebook: Click speech of likes, comments and shares," *Proc of the International Conference on Digital Information Management*, IEEE, Sep. 2017, pp. 257-262.
- [14] A. Y. Chua, and S. Banerjee, "Rumors and rumor corrections on Twitter: Studying message characteristics and opinion leadership," *Proc of the International Conference on Information Management*, IEEE, May 2018, pp. 210-214.
- [15] L. Bode, and E. K. Vraga, "See something, say something: Correction of global health misinformation on social media," *Health Communication*, vol. 33, no. 9, pp. 1131-1140, 2018.
- [16] S. Lewandowsky, U. K. Ecker, C. M. Seifert, N. Schwarz, and J. Cook, "Misinformation and its correction continued influence and successful debiasing," *Psychological Science in the Public Interest*, vol. 13, no. 3, pp. 106-131, 2012.
- [17] B. Nyhan, and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. (2), pp. 303-330, 2010.
- [18] S. Lewandowsky, W. G. Stritzke, A. M. Freund, K. Oberauer, and J. I. Krueger, "Misinformation, disinformation, and violent conflict: From Iraq and the 'War on Terror' to future threats to peace," *American Psychologist*, vol. 68, no. 7, pp. 487-501, 2013.
- [19] A. Pal, A. Y. K. Chua, and D. H. L. Goh, "Salient beliefs about sharing rumor denials on the Internet," *Proc of the International Conference on Ubiquitous Information Management and Communication*, ACM, Jan. 2018, p. 57.
- [20] L. Rösner, and N. C. Krämer, "Verbal venting in the social web: Effects of anonymity and group norms on aggressive language use in online comments," *Social Media + Society*, pp. 1-13, 2016.
- [21] V. S. Chavan, and S. Shylaja, "Machine learning approach for detection of cyber-aggressive comments by peers on social media network," *Proc of the International Conference on Advances in Computing, Communications and Informatics*, Aug. 2015, pp. 2354-2358.
- [22] R. Young, S. Miles, and S. Alhabash, "Attacks by Anons: A content analysis of aggressive posts, victim responses, and bystander interventions on a social media site," *Social Media + Society*, pp. 1-14, 2018.
- [23] K. Coe, K. Kenski, and S. A. Rains, "Online and uncivil? Patterns and determinants of incivility in newspaper website comments," *Journal of Communication*, vol. 64, pp. 658-679, 2014.
- [24] J. Gonçalves, "Aggression in news comments: How context and article topic shape user-generated content," *Journal of Applied Communication Research*, vol. 46 no. 5, pp. 604-620, 2018.
- [25] A. A. Anderson, S. K. Yeo, D. Brossard, D. A. Scheufele, and M. A. Xenos, "Toxic talk: How online incivility can undermine perceptions of media," *International Journal of Public Opinion Research*, vol. 30, no. 1, pp. 156-168, 2018.
- [26] I. Rowe, "Civility 2.0: A comparative analysis of incivility in online political discussion," *Information, Communication & Society*, vol. 18, no. 2, pp. 121-138, 2015.
- [27] A. Santoso, "Polisi Usut Netizen yang Sebar Hoax Bocah Perusuh Tewas Dipukuli di Kampung Bali," *Detik.com.*, May 25, 2019, Retrieved from <https://news.detik.com/berita/d-4564462/polisi-usut-netizen-yang-sebar-hoax-bocah-tewas-dipukuli-di-kampung-bali>
- [28] N. Habibie, "Pengakuan Andri Bibir, Pria yang Dipukul Brimob di Kampung Bali," *Liputan6.com.*, May 25, 2019. Retrieved from <https://www.liputan6.com/news/read/3975513/pengakuan-andri-bibir-pria-yang-dipukul-brimob-di-kampung-bali>
- [29] Z. Lilland, "Viral Foto 'Brimob Sipit', Polri: Semua Personel WNI," *Detik.com.*, May 22, 2019. Retrieved from <https://news.detik.com/berita/d-4560137/viral-foto-brimob-sipit-polri-semua-personel-wni>
- [30] A. Santoso, "Anggota Brimob yang Dituduh dari China Buka Suara: Saya Asli Indonesia", *Detik.com.*, May 24, 2019. Retrieved from <https://news.detik.com/berita/d-4563679/anggota-brimob-yang-dituduh-dari-china-buka-suara-saya-asli-indonesia>
- [31] D. Halim, "Penyebar Hoaks Ada Brimob dari China Minta Maaf," *Kompas*. May 24, 2019. Retrieved from <https://nasional.kompas.com/read/2019/05/24/18241461/penyebar-hoaks-ada-brimob-dari-china-minta-maaf>
- [32] Katadata, "YouTube, Medsos No. 1 di Indonesia," Mar. 6, 2019. <https://katadata.co.id/infografik/2019/03/06/youtube-medsos-no-1-di-indonesia>